

1 Formal Prolegomena

First, some notational definitions. We have tabulated them below.

- (\mathcal{U}) The set of universities and colleges in question.
- (\mathcal{D}) Financial space. In general, this is the space of donations – this might have multiple dimensions over \mathbb{R} , depending on the specific categories of money we’re interested in. It might also contain certain categories of expenditures that we might be able to influence by placing constraints on how institutions are allowed to spend their money.
- (\mathcal{T}) The space of times for which we have data.
- (\mathcal{V}) The vector space of student metric variables $\{v_i\}$. Note that at this point, we have not yet committed ourselves to any such choice of variables, and so \mathcal{V} includes also negative and neutral indicators of success.
- ($d\mathcal{V}$) In a continuous setting, this would be the \mathcal{V} differential 1-form, but here it is simply a running difference over time.

We will also be interested in a “sliding window” of times trailing a given time; if we’re interested in a subspace of time $T \subset \mathcal{T}$, we can construct this window as $\mathcal{W}_T : T \rightarrow \mathcal{D}$. Note that in the case of a discrete T , comprised of n time steps, $\mathcal{W}_T = \mathcal{D}^n$. All of our data is discrete, of course, but this construction also works for continuous subspaces of \mathcal{T} , which means we can think of different choices of T as appropriate approximations, instead of sets with fundamentally different structures.

With this framework, we can now formulate the problem more precisely. To do any kind of induction at all, it is necessary to make some commonplace but sometimes very wrong independence assumptions. Here’s our first and most central one: we will assume that the effectiveness with which an institution can use money does not change over time¹. We can now talk about the effect of donor money over time on the variables in \mathcal{V} as a mapping

$$F : \mathcal{W}_T \rightarrow d\mathcal{V} \tag{1}$$

That is – a function which takes the financial input we control, and changes student metric variables in some way. This is to be distinguished from a function $\phi : \mathcal{W}_T \rightarrow \mathcal{V}$ that estimates The first goal of this analysis is to learn this function.

Supposing that we have a suitably robust approximation for F , we still

1. Apply Machine Learning techniques to estimate the effect of donations on the

¹This is a reasonable assumption to make; while technically invalid, it seems very natural to judge an institution by its past performance – indeed, this is the best we can hope for from a dataset